The Evolving Relationship Between Public Sentiment and Bitcoin Market Volatility

Jimmy Zhou AP Research
Prince of Peace Christian School
Carrollton, Texas
Apr. 17, 2025
yunchengzhou@gmail.com

***Abstract*—This study researches the relationship between public sentiment on the subreddit r/Bitcoin and Bitcoin market volatility. Utilizing the sentiment of posts determined using VADER and market volatility calculated via Parkinson's formula, Pearson and Spearman correlational analysis revealed a weak negative correlation between sentiment and volatility. Moreover, Granger causality tests demonstrated that while Bitcoin volatility significant influences Reddit entiment, sentiment does not have any predictive capabilities on volatility.**

***Index Terms*—Bitcoin, Cryptocurrency, Reddit, Sentiment Analysis, Market Volatility, Granger Causality, VADER, Social Media Analytics, Time Series Analysis**

I Introduction

Bitcoin, created in 2009 by an anonymous online persona known only as Satoshi Nakamoto, has become one of the most widely recognized and traded global assets. The creation of Bitcoin was the first ever implementation of a cryptocurrency. Cryptocurrency is a form of digital currency that operates without a central authority like a bank or government. Instead, a cryptocurrency is controlled by a decentralized network of computers, which records and verifies transactions over a blockchain, a type of digital ledger, to make sure systems are as secure as possible. In recent years, the cryptocurrency market has played a key role in advancing the global financial market. The "global cryptocurrency market cap" has reached "2.63 trillion US Dollars, a 50-fold increase when compared with the May 2017 market cap of 50 billion US Dollars" [1]. Additionally, "the term Bitcoin ranked as the ninth trendiest search in the US" [2]. Interestingly, unlike traditional financial markets, the value of cryptocurrency is heavily influenced by the public's opinions, especially on social media platforms. Modern social media serves as a space to exchange information, and it allows for the formation of diverse public opinions. These platforms, such as Twitter, Reddit, and Telegram, have become extremely influential in the crypto-trading community; one source states that "social media platforms and online forums play a vital role in shaping the market dynamics of cryptocurrencies" [3]. By understanding the meaning behind posts shared on social media, data scientists and financial analysts will be able to understand the true influence social media has on cryptocurrency. Therefore, because the value of cryptocurrency is heavily influenced by public opinion, this study hopes to understand whether or not cryptocurrency price is influenced by online discussions on social media sites. By analyzing this relationship, this study aims to contribute to the body of knowledge surrounding social media and cryptocurrency.

II Literature Review

Reviewing current research regarding the broader body of knowledge surrounding this study has played a huge role in the development of the methodology used in this research study. Additionally, upon analyzing current baseline studies, a clear gap has been identified in this paper.

In order to analyze the implied meaning behind a piece of text, recent studies have relied on the use of sentiment analysis. One study performed a time-series analysis done on the relationship between the cryptocurrencies Bitcoin and Ethereum and sentiment collected from 24 subreddit communities on the social media site Reddit [4]. This study collected data from Reddit "due to the longer character limit for posts and availability of the data" [4]. Additionally, the sentiment analysis tool used in this study was VADER, found in the Python NLTK library. According to a study done over sentiment analysis specifically on social media related text, "the application of VADER sentiment analysis on the dataset of paper titles demonstrated its effectiveness in capturing and visualizing sentiment trends"[5]. This studied engineerd 112 time-series features from the sentiment data and performed bivariate Granger causality tests

against Bitcoin and Ethereum closing price; additionally, the study also built classification models to try and predict daily cryptocurrency price movements alongside lagged price data. Upon doing so, the study found that ”not only are the Reddit-derived features correlated with price changes, but they contain information that improve market predictions,” and found that there was a strong correlation between sentiment and Bitcoin and Ethereum price movements [4]. Another study conducted a comparative study evaluating different large language models (LLM) and natural language processing models (NLP) for cryptocurrency [6]. The researchers fine-tuned three LLMs - GPT-4, BERT, and FinBERT - on a dataset of 31,037 cryptocurruency news articles scraped from Cryptonews and Cointelegraph. Fine-tuning is when a pre-trained model is trained on an additional dataset to improve its performance in a more specific field. In this case, the additional dataset is the sentiment data and cryptocurrency price data. In this study, the sentiment was extracted using TextBlob, another sentiment analysis model. After this training, the data was tested against a benchmark, and the researchers found that ”the models demonstrated almost flawless prediction accuracy post-fine-tuning with few-shot learning, achieving a maximum of 86.7%” [6]. A different study was done determining the correlation between the sentiment of both Reddit and Twitter data and Bitcoin prices . This study gathered Reddit data from the Reddit API using pushshift.io, and it gathered Twitter data from the Twitter API. It then used VADER to analyze the overall sentiment of the posts and comments, and Bitcoin price data was scraped from Coinbase’s historical Bitcoin database. The researchers then used Pearson’s correlation coefficient to determine the correlational relationship between sentiment scores and Bitcoin prices. Additionally, another study was done determining the correlation between the sentiment of both Reddit and Twitter data and Bitcoin prices [7]. This study gathered data from the Reddit API, using pushshift.io, and from the Twitter API. It then used VADER to analyze the overall sentiment for the posts and comments, and Bitcoin price data was scraped from Coinbase. The study confirmed existing research suggesting a relationship between social media sentiment and Bitcoin price and stated that ”there is some correlation between social media sentiment and Bitcoin value” [7].

However, a few studies found that there was no significant correlation between sentiment and cryptocurrency. One study extracted Twitter posts using the Twitter API and found the sentiment of the posts using both TextBlob and VADER [8].Upon analyzing the sentiment, the study found ”an overall increasing trend in polarity with several significant peaks,” but this trend was not enough to provide any significant results by itself [8]. The researchers then collected historical Bitcoin prices from CoinGecko’s API across a period of 1481 days. After comparison between these two variables, the study calculated a Pearson correlation coefficient of 0.0158, which is considered extremely low and shows little to no correlation. Therefore, as stated in the study, ”there is no correlation between daily Twitter sentiment and Bitcoin price movements” [8]. Similarly, a different study analyzed cryptocurrency prices with Twitter sentiment and found that ”the correlation achieved was close to 0, showing us the lack of correlation between cryptocurrency price and sentiment” [9]. This study used VADER to extract sentiment from Twitter posts and collected Bitcoin, Ethereum, and Chainlink price data from a historical dataset. The researchers then calculated 3 cross-correlation coefficients to determine the relationship between the two variables - Pearson’s, Kendall’s, and Spearman’s - but because the coefficients were too small, the researchers were unable to determine a statistically significant relationship. Likewise, a predictive study analyzed approximately 567 thousand tweets and collected price data from the top 12 cryptocurrencies over a period of 1255 days [10]. The researchers then used DistilBERT, a distilled version of an LLM (BERT) by HuggingFace, to calculate sentiment along with TextBlob and VADER to determine subjectivity and polarity. The study then created several prediction models to predict future cryptocurrency prices. After doing so, the study states that upon comparing results to the test dataset, the prediction models were ”not successful in reaching either the baseline accuracy or returns using it” [10].

Through the analysis of current literature related to the body of knowledge surrounding online sentiment and its relationship to cryptocurrency, a clear conclusion can be drawn. Most, if not all, of the studies analyzed using Twitter sentiment found that there was not a strong correlation with cryptocurrency prices. On the other hand, studies done on different types of social media platforms and news sources seemed to display a more significant correlation between the two variables. However, the studies that did involve the analysis of Reddit sentiment only discussed the correlation between Reddit sentiment and cryptocurrency open and closing price on a day-to-day basis, and most of these studies focused on several different large cryptocurrency-based subreddits. This demonstrates a clear gap in the literature, so this study will discuss the correlation between the sentiment of social media posts found specifically on a singular Reddit community and market volatility. Additionally, this study will analyze the cryptocurrency Bitcoin because it is the largest and thus the most influential cryptocurrency. By demonstrating a clear focus on the volatility of Bitcoin, this study hopes to provide a more in-depth understanding

of social media's influence on Bitcoin market behavior. This creates the research question:

How does public sentiment regarding the cryptocurrency Bitcoin on the social media site Reddit correlate with, or potentially influence, Bitcoin market volatility over time?

After analyzing the current literature in this field of study, I believe that there will be a statistically significant correlation between the sentiment of Reddit data and the market volatility of Bitcoin. Furthermore, I hypothesize that sentiment will have a statistically significant influence on market volatility, and market volatility will have a statistically significant impact on sentiment.

III Methodology

A. Python Modules

The data analysis in this study was performed using the coding language Python, which is commonly used by data scientists to analyze large sets of data. Additionally, several Python libraries were used for comprehensive data analysis: Pandas, Matplotlib, Seaborn, and NLTK. Pandas is a "fast, powerful, flexible and easy to use open source data analysis and manipulation tool, built on top of the Python programming language" and it serves as a foundational data manipulation tool with its introduction of DataFrames [11]. Matplotlib and Seaborn both provide important data visualization capabilities to create important graphs and tables. Finally, NLTK (Natural Language Tool Kit) provides the VADER sentiment analysis tool and aids in data preprocessing through its use of stop words and tokenization. The code used in this project is found here https://github.com/Jimboni-tech/APResearchProj.

B. Data Collection

This study utilizes two primary datasets: Reddit posts and historical Bitcoin price data. The Reddit data was collected from the r/Bitcoin subreddit, focusing on posts made between December 31, 2023, and December 31, 2024.

This study accessed a database containing data collected, extracted, split, and re-packaged by u/raiderbdev and u/Watchful1. This database collected its data using pushshift.io, a big-data collection model created by Jason Baumgartner (known online as u/Stuck-In-The-Matrix); pushshift.io was chosen to collect data for this database because of its ability to obtain data directly from Reddit's official API and because of its ability to store extremely large quantities of historical data.

Upon accessing this database, the downloaded data was stored in a torrent file due to its sheer size. A torrent file is a peer-to-peer file-sharing method that breaks data into small parts for efficient downloading. Torrents are extremely valuable for academic research because they allow for access to large datasets without having to rely on a single server.

C. Data Processing

The dataset analyzed in this study focused specifically on r/Bitcoin, the largest Bitcoin-related subreddit. The raw data was initially stored in a ZStandard-compressed NDJson format, a high-performance data compression standard designed for efficiency. A custom Python script was implemented to decompress the dataset and filter the contents based on the designated time frame (December 31, 2023, to December 31, 2024). The script systematically extracted and parsed all Reddit submissions from this time period. The extracted data was then reformatted into a structured CSV file. This study extracted the key metadata points from the dataset:

- Author: The Reddit username of the post or comment creator.
- Created-UTC: The timestamp of when the content was posted.
- Title: The headline of a Reddit post.
- Self-text/Body: The textual content of submissions.

The self-text section of the DataFrame needed to be preprocessed before sentiment analysis could be applied to it. One study states that "it is necessary to pre-process raw text to ensure that sentiment analysis will focus on the actual and meaningful content" [12]. Therefore, this study will use a custom-made preprocessing pipeline to remove:

- URLS/Markdown Links
- Special Characters - Symbols and emojis (eg., "&", "#")
- Stop Words - Common words (e.g., "the", "and") that do not add meaningful content
- Empty Submissions

This study used a list of stop words from the NLTK Python library. After data processing, the dataset retained 39.61% of submissions, or 35111 of the original 88631. While this dataset is much smaller than before, removing unnecessary entries is essential for accurate use of sentiment analysis.

D. Sentiment Analysis

After careful analyzation of modern literature related to the topic of this study, this study will utilize the VADER (Valence Aware Dictionary for Sentiment Reasoning) NLTK (Natural Language Toolkit) module. VADER is "designed to perform well on social media text but

effective across other text forms as well" [5]. Evidently, VADER would be the most optimal sentiment analysis tool for the Reddit dataset. VADER is a lexicon- and rule-based sentiment analysis tool specifically attuned to informal language, such as slang and emoticons. Its lexicon contains a list of words, emojis, and acronyms pre-scored for sentiment intensity based on real human raters. To compute sentiment scores, VADER applies five heuristic rules:

- polarity shifts (e.g., "not good")
- modulation from degree modifiers (e.g., "very" amplifies, "slightly" dampens)
- contrasting effects from conjunctions like "but"
- capitalization emphasis (e.g., "GREAT")
- sentiment decay with the distance between terms

Upon inputting text, the tool synthesizes these features into four different sub-categories: positive, neutral, negative, and compound. The compound metric reflects overall sentiment and ranges from [-1,1], while positive, neutral, and negative values represent proportional sentiment distributions; these values range from [0,1]. For example, upon entering the phrase, "VADER is the best sentiment analysis tool," the output would yield a negative value of '0.0,' a neutral value of '0.572,' a positive value of '0.428,' and a compound value of '0.6696'. This study will focus only on the compound values of each post, as it provides the most information regarding the overall meaning of a text. This study will first store the Reddit data inside a Pandas DataFrame consisting of two columns, created-UTC and self-text. Next, the program will standardize the created-UTC, in Year/Month/Day 00:00:00 format, to a standard date format, Year-Month-Day. The program will then import NLTK and initialize the Sentiment Intensity Analyzer from the VADER module. This analyzer will then be applied to all pieces of data in the self-text column of the Pandas DataFrame to calculate the sentiment scores of each post's primary text, and the corresponding compound score of each piece of text will be added to a new column within the DataFrame. The sentiments are then grouped by date, and the average of the compound scores of each post spread across a certain date is calculated and concatenated into another DataFrame. This new DataFrame has two columns, a date and its corresponding average sentiment score.

Date	Sentiment Score
2023-12-31	0.183
2024-01-01	0.109
2024-01-02	0.105
2024-01-03	0.056
2024-01-04	0.105
2024-01-05	0.127
2024-01-06	0.112
2024-01-07	0.121
2024-01-08	0.098
2024-01-09	0.088
2024-01-10	0.120
2024-01-11	0.104
2024-01-12	0.089
2024-01-13	0.107
2024-01-14	0.114

Table I. This table depicts 15 days of sample data spanning from December 31, 2023 to January 14, 2024.

E. Bitcoin Data Collection and Processing

The Bitcoin data in this study was taken from CoinMarketCap's historical Bitcoin dataset. After downloading the data, this study used a simple Python script to extract the date, open price, high price, low price, close price, volume, and market cap of every single day across the 365 day time period. This data was first stored in a Pandas DataFrame consisting of 7 columns corresponding with the 7 pieces of data collected from each day. The script then changed the date from UTC formatting to the standard date format. Next, for each day, Parkinson's historical volatility formula was applied to each row.

$$\sigma = \frac{1}{\sqrt{4\ln(2)}} \times \frac{\text{High} - \text{Low}}{\text{Close}} \tag{1}$$

The calculated data was then concatenated into another DataFrame with the columns Date and Market Volatility for further analysis.

F. Combined Analysis

To investigate the relationship between Bitcoin market volatility and Reddit sentiment, a multi-step data analysis process was conducted. First, a DataFrame was created by merging the Bitcoin DataFrame and the Reddit sentiment DataFrame. In this new DataFrame, each date and its corresponding average sentiment and market volatility value were grouped together so that the new DataFrame only had the columns date, market volatility, and average sentiment.

Date	Volatility	Sentiment
2024-01-02	0.022975	0.378937
2024-01-03	0.065323	0.280046
2024-01-04	0.028780	0.364737
2024-01-05	0.021624	0.349235
2024-01-06	0.010306	0.335889
2024-01-07	0.011354	0.377040
2024-01-08	0.052798	0.295889
2024-01-09	0.034171	0.320261
2024-01-10	0.041267	0.357989
2024-01-11	0.041778	0.293123

Table II. This table depicts the merged DataFrame with 10 days of sample data spanning from January 2, 2024, to January 11, 2024

In order to evaluate the linear and rank-based relationships between sentiment and volatility, both the Pearson correlation coefficient and the Spearman rank correlation coefficient were computed. The Pearson correlation measures the strength and direction of a linear relationship, while the Spearman correlation assesses the monotonic relationship between the variables, making it more robust to non-linear patterns. Pearson's correlation coefficient can be calculated with the formula:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{2}$$

Spearman's correlation coefficient can be calculated with the formula:

$$\rho = 1 - \frac{6\sum d_i^2}{n(n^2 - 1)} \tag{3}$$

Next, to explore potential predictive relationships between sentiment and volatility, Granger's causality test was conducted. The test was performed to test whether sentiment scores Granger-caused Bitcoin market volatility, and whether market volatility Granger-caused sentiment scores. For both directions of causality, a lag from one to seven days was used to capture short-term predictive patterns. The formulas for the two models of the Granger causality test are:

$$Y_t = \alpha_0 + \sum_{i=1}^{p} \alpha_i Y_{t-i} + \epsilon_t \tag{4}$$

$$Y_t = \beta_0 + \sum_{i=1}^{p} \beta_i Y_{t-i} + \sum_{j=1}^{p} \gamma_j X_{t-j} + u_t \tag{5}$$

IV Results

This study collected a total of 53,520 entries of data across a total of 365 days. The sentiment ranged from 0.10 to 0.491, with a mean sentiment of 0.328, a standard deviation of 0.059, and a skewness of -0.219. This skewness value means that the data is relatively symmetrical, and the standard deviation means that the sentiment is not spread very far from the mean.

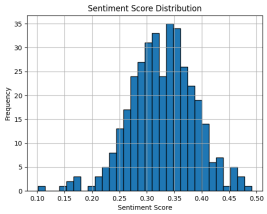

Figure 1. Distribution of Sentiment

Additionally, upon graphing sentiment score on the x-axis and Parkinson's volatility on the y-axis, a regression line can be formed. Upon review, a slight negative correlation can be seen with a general clustering of points around the regression line, so some sort of correlation can be seen in this graph.

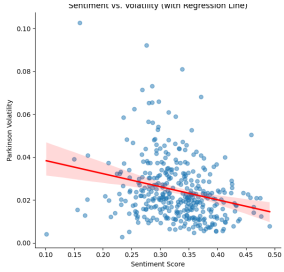

Figure 2. Regression Line

More information must be calculated to form a more accurate and reliable conclusion, however. Therefore, this study will make use of Pearson's correlation coefficient and Spearman's correlation coefficient. These coefficients were calculated using the data points located in the sentiment and market volatility columns of the Pandas DataFrame. Upon calculation, a Pearson's correlation value of -0.252 and a Spearman's correlation value of -0.274 were outputted. Based on Pearson's correlation, it can be inferred that as sentiment becomes more and more positive, volatility tends to slightly decrease. On the other hand, Spearman's correlation demonstrates that higher sentiment has a slight correlation with lower volatility. These findings suggest a weak negative relationship between sentiment and market volatility. Because there are several limitations regarding these correlational measures (the correlation is not strong enough to suggest predictive power, volatility may have a delayed response to sentiment), a time-series dual-axis plot can be used to observe causation.

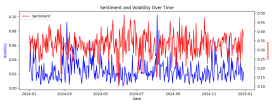

Figure 3. Time Series Dual Analysis

As seen in figure 5, there are several divergences and similarities between the two variables. For example, there are several moments where the volatility and sentiment are complete opposites of each other (one moves up while the other moves down), and there are other moments where it seems like one variable is following the other. Because there is mixed feedback regarding this causation, further analysis must be done to come to a reasonable conclusion. Therefore, this study will implement the Granger causality test to further explore the direct relationship between the two variables. The Granger causality test is a statistical hypothesis test used to determine whether one time series has the ability to predict another. The Granger causality test is perfect for this study because it specifically evaluates whether past values of one variable contain enough information to help predict future values of another variable. This test operates in two models: one in which the target variable is regressed on only its own lagged values, and another in which the target variable is regressed on its own lagged values and the lagged value of the predictor variable [13]. If the inclusion of the lagged predictor variable has a large impact on the prediction of volatility, then that variable has a "Granger-cause" on the target variable.

Upon importing the Granger causality test from the statsmodel Python module, the Granger causality test was used to calculate the influence of sentiment on volatility and the influence of volatility on sentiment with a max lag of 7 days.

Table III. Granger Causality Test Results (Lag = 3)

Test Type	**Statistic**	**p-value**
SSR F Test	F = 1.5167 (df = 3, 354)	0.2099
SSR Chi-squared Test	χ^2 = 4.6400 (df = 3)	0.2001
Likelihood Ratio	χ^2 = 4.6104 (df = 3)	0.2027
Parameter F Test	F = 1.5167 (df = 3, 354)	0.2099

Table III represents one result of the Granger causality test on the influence of sentiment on volatility at lag = 3 (whether sentiment data from the past 3 days helps to predict volatility today). In Table IV, it can be seen that the p-values are well above 0.05. In this case, the null hypothesis would be that sentiment does not Granger-cause volatility. Because the p-value is so high, we do not reject the null hypothesis, so there is no statistically significant evidence that sentiment causes volatility.

Lags	p-value (F-test)
1	0.8992
2	0.6495
3	0.2099
4	0.3618
5	0.5135
6	0.4127
7	0.4007

Table IV. Influence of Sentiment on Volatility

Lags	p-value (F-test)
1	0.0171
2	0.0378
3	0.0347
4	0.0633
5	0.0981
6	0.1221
7	0.0964

Table V. Influence of Volatility on Sentiment

In Table V, the null hypothesis would be that volatility does not Granger-cause sentiment. Because the p-values in lags 1-3 are below 0.05, for lags 1-3, we reject the null hypothesis. Therefore, in the range of lag 1-3, there is statistically significant evidence that volatility can help predict sentiment.

V Conclusion and Discussion

In conclusion, although there was a slight negative correlation between Reddit sentiment and Bitcoin volatility, there is not enough evidence to suggest sentiment has a strong predictive influence on volatility; however, volatility seems to have a statistical influence on sentiment, as predicted. These findings are similar to other studies analyzing Reddit sentiment, however, the understanding of sentiment having a strong predictive influence on volatility is a newer understanding [4][7].

A. Implications

Because the Granger causality test determined that market volatility does Granger-cause sentiment, it can be understood that Reddit users have a high reaction to volatility. Additionally, sentiment should be seen as a real-time stress indicator instead of a volatility predictor. While there is a slight correlation between the two, users should not use the sentiment of online discussions in the subreddit r/Bitcoin to try and predict future Bitcoin price movements. While high volatility does in fact lead to negative sentiment, the same can not be said of the opposite; sentiment does not lead to a change in Bitcoin volatility.

Additionally, this study aids in new understandings regarding trading risk and management. Because there is only a slight negative correlation between the two, there is no actionable sign, and traders and investors should not invest just because of changes in sentiment. After all, correlation does not equal causation.

B. Limitations

This study has many notable limitations. For one, VADER is not 100% accurate. While VADER is quite an effective sentiment analysis tool, especially on social media posts, it can misinterpret sarcasm and slang. Moreover, there are many spam posts, memes, and bots on Reddit that may have a heavy impact on the overall average sentiment data. If these factors were resolved, there may be an even clearer understanding of the relationship between sentiment and Bitcoin volatility.

Additionally, external factors may have an influence on volatility. Changes such as political events, economic trends, or influential figures may have a large impact on the Bitcoin market, and these factors may not be captured by Reddit sentiment. Therefore, the relationship between sentiment and market volatility may be confounded, and the observed results of this study may come from a factor outside of online sentiment.

Finally, the largest limitation of this study was the size of the dataset used. The machine used for this study, a MacBook Air with an M3 chip, simply could not handle the terabyte of data stored in the database. If this study had access to online resources such as RunPod, the size of the dataset could grow exponentially, and the resulting data may look very different.

C. Future Research

This study leads to many innovations in future research. This study found a gap in current literature and decided to compare market volatility to Reddit sentiment, but future research could go a step further by training LLMs or predictive regression models on Bitcoin market volatility and online sentiment to determine whether sentiment truly does not influence market volatility. Additionally, future studies could implement transformer-based models like BERT, RoBERTa, or FinBert specifically trained on crypto subreddits to better extract sentiment. By using these models, a better understanding of the posts on Reddit can be discovered, potentially leading to a new understanding of the relationship between sentiment and market volatility. Finally, future studies could categorize Reddit posts by topic and analyze whether sentiment relating to specific aspects of Bitcoin (eg: technical, financial, investing) has stronger correlations with volatility.

References

[1] R. B. Hurjui, "The effect of market sentiment on cryptocurrency prices," Master's thesis, Erasmus School of Economics, M.S. thesis, Erasmus University Rotterdam, 2022.

[2] L. A. L. Camou, "Reddit as a prediction tool for crypto-assets," *Revista Brasileira de Finanças*, vol. 20, no. 1, 2022, Accessed: Apr. 2025. [Online]. Available: https://link.gale.com/apps/doc/A702395646/AONE.

[3] G. Vlahavas and A. Vakali, "Dynamics between bitcoin market trends and social media activity," *Preprints*, 2024, Preprint. DOI: 10.20944/preprints202407.0112.v1.

[4] S. Wooley, A. Edmonds, A. Bagavathi, and S. Krishnan, "Extracting cryptocurrency price movements from the reddit network sentiment," in *Proceedings of the [Conference Name]*, 2023.

[5] D. C. Youvan, *Vader sentiment analysis tool documentation*, Accessed: Apr. 2025, 2024.

[6] K. I. Roumeliotis *et al.*, "Llms and nlp models in cryptocurrency sentiment analysis: A comparative classification study," *Big Data and Cognitive Computing*, vol. 8, no. 6, Jun. 2024. DOI: 10.3390/bdcc8060063.

[7] A. Salač, "Forecasting of the cryptocurrency market through social media sentiment analysis," M.S. thesis, University of Twente, 2023.

[8] K. Byc and S.-C. Ilinca, "The relationship between social media sentiment and bitcoin price volatility," M.S. thesis, Copenhagen Business School, 2021.

[9] H. A. Kang, "Correlating cryptocurrency price with twitter tweet sentiment," M.S. thesis, M.S. thesis, University of North Carolina at Chapel Hill, Aug. 2020.

[10] V. Koltun and I. P. Yamshchikov, "Pump it: Twitter sentiment analysis for cryptocurrency price prediction," *Risks*, vol. 11, no. 9, 2023. DOI: 10.3390/risks11090159.

[11] Pandas Development Team, *Pandas: Powerful python data analysis toolkit*, https://pandas.pydata.org/, Version 2.1.4, accessed Apr. 15, 2024, n.d.

[12] A. Zhumagaziyev, "Social media sentiment and cryptocurrency performance," Unpublished manuscript, 2023.

[13] X. Song and A. Taamouti, "A better understanding of granger causality analysis: A big data environ-

ment," *Oxford Bulletin of Economics and Statistics*, vol. 81, pp. 911–936, 2019. DOI: 10.1111/obes.12288.